

Non-native β -sheet formation: insights into protein amyloidosis

Chinlin Guo^{†‡}, Herbert Levine[‡], and David A. Kessler^{*}

[†]*Department of Molecular Cell Biology,
Harvard University, 16 Divinity Avenue,
Room 3007, Cambridge, MA 02138*

[‡]*Department of Physics, University of California,
San Diego 9500 Gilman Drive, La Jolla, CA 92093-0319 and*

^{*}*Department of Physics, Bar-Ilan University, Ramat-Gan, Israel*

Abstract

Protein amyloidosis is a cytopathological process characterized by the formation of highly β -sheet-rich fibrils. How this process occurs and how to prevent/treat the associated diseases are not completely understood. Here, we carry out a theoretical investigation of sequence-independent β -sheet formation, based on recent findings regarding the cooperativity of hydrogen-bond network formation. Our results strongly suggest that *in vivo* β -sheet aggregation is induced by inter-sheet stacking dynamics. This leads to a prediction for the minimal length of susceptible polymer needed to form such an aggregate. Remarkably, the prediction corresponds quite well with the critical lengths detected in poly-glutamine-related diseases. Our work therefore provides a theoretical framework capable of understanding the underlying mechanism and shedding light on therapy strategies of protein amyloidosis.

Protein amyloidosis [1] is seen in the prion-related transmissible spongiform encephalopathies such as Creutzfeldt-Jakob disease [2] and nine fatal poly-glutamine (poly-Q) related neurodegenerative disorders including Huntington disease, Kennedy disease, six spinocerebellar ataxias, and dentatorubral pallidoluysian atrophy [3]. It has also been found that normal peptides can undergo misfolding to form amyloid aggregate [4]. Thus, protein amyloidosis can generally involve a sequence-independent mechanism; the critical questions then would be how this process is initiated and how to prevent its onset.

Although many different proteins can form amyloidosis [4], it has been shown that the agents in aggregation-prone peptides are their N-terminal oligopeptide repeats, which have a tendency to form non-native β -sheets that in turn can serve as templates to aggregate additional mis/un-folded peptides into elongated fibrils [3, 5]. The rate-limiting step in this process is the nucleation of the non-native β -sheets [6]; moreover, the aggregation tendency correlates well with the length of repeats Δl [3, 5]. For instance, in the Huntington disease gene, the “normal” range of poly-Q repeats is 6 to 34 CAGs, and the pathological forms involve tracts of 37 or longer repeats [3]. Likewise, a deletion of prion N-terminal repeats can suppress its spontaneous aggregation rate [5].

What is the biological implication of these facts? In principle, a longer Δl implies a larger non-native β -sheet formation. Under normal physiological conditions, the nucleation of non-native β -sheets is suppressed by intracellular scavengers, chaperones and the ubiquitin-proteasome system which detect and convert/degrade misfolded peptide intermediates [7, 8]. It has however been found that when Δl exceeds the pathological threshold, scavengers do not work; instead, they become non-functionally engaged with the non-native structure [9]. This impairment of molecular scavengers indicates that the formation of large non-native β -sheets can undergo a rapid “two-state-like” transition, i.e., proceeding without any intermediate of long enough detectable lifetime. It also implies that the free energy barrier ΔG separating the “two-states” is too high to be overcome by intracellular machinery utilizing regular energy sources (ATP, e.g.). Thus, elucidating how the repeat length Δl is related to these two properties is the main goal of present work.

Methods

We study the sequence-independent thermodynamics and kinetics of an extensive homopolymer (to mimic the oligopeptide repeats) that is either at the tail or an inserted loop of

a pathological peptide, Fig. 1. We are interested in how this homopolymer can form ordered β -sheets that provide the most advantageous fibril template (of course, the homopolymer can form single-strand alpha-helix and multi-strand amorphous aggregate, which nevertheless is less pathogenic and is not our major concern). In particular, we identify the homopolymer length Δl and corresponding β -sheet topology at which the two-state-like behavior emerges with a large free energy barrier ΔG . Furthermore, we investigate the predominant kinetics and patterns (Fig. 1) characterizing the transition; this knowledge can then be applied to therapy issues such as drug design targets and expected dose-response curves.

Our study proceeds in two stages. First, we study the formation of a single-layer β -sheet (from one isolated pathological peptide) consisting of $M+1$ strands each of length L residues, making use of a model incorporating the cooperativity inherent in H-bond formation. This cooperativity has its microscopic origin in the collective expelling of water molecules [10] from inter-chain positions in the nascent sheet [11] that results in the free energy of formation of a single H-bond being dependent on the state of the neighboring bonds along the hairpin axis (intra-hairpin coupling) and between different hairpin segments along the “fibril” elongation axis (inter-hairpin coupling) [11, 12]. The thermodynamics is then calculated by Monte Carlo enumeration of the partition function [12]. Furthermore, we utilize the idea of droplet formation [13, 14, 15] to analyze the dominant kinetic pathway for the folding transition. Afterwards, we incorporated sheet-sheet interactions (“stacking”) and made a simple assumption that the free energy change due to a single H-bond formation within one peptide has a linear dependence on the local H-bond density contributed by other nearby sheets. Using a mean-field approach, we then can compute the H-bond density self-consistently. The precise form of our Hamiltonian as well as the details of our calculations are available in the appendix and supplement.

Results and Discussion

Two-state single-layer β -sheet formation. Results for this case are shown in Fig. 2. First, there is indeed a region of two-state thermodynamics, bounded by two regions inside which the formation of non-native β -sheets always encounters intermediate states (and hence can be eliminated by intracellular scavengers), Fig. 2a. This implies, in accord with experimental observation [16], that too long peptide N-terminal repeats do not aggregate *in vivo*. Second, the kinetic barrier ΔG for two-state transition is large (≥ 25 kcal/mol for

number of H-bond $\geq 10^2$, Fig. 3a, e.g.), meaning that β -sheet formation in two-state region cannot be easily reversed by intracellular scavengers. Inside the two-state region, we also observe two distinct patterns that dominate the kinetics, as a function of the topology.

Mechanism facilitating *in vivo* amyloidosis. The large ΔG also implies that the transition would not occur in any reasonable time span. Thus, there must be another mechanism to enhance non-native β -sheet formation. The most likely one is the inter-sheet stacking, as we note that folded β -sheets tend to stack into a nearly anhydrous structure even when rich with polar side-chains [17]. This is probably because a) hydration of a large-scale 2-dimensional structure requires extensive, homogeneously distributed H-bond donors and acceptors to interact with water molecules, and b) the formation of extensive β -sheet H-bond network consumes all available backbone electron pairs and protons (including the $C_\alpha H \cdots OC$ interaction [11, 18]). Eventually, neighboring sheets prefer squeezing out inter-sheet water molecules. This leads to a synergistic H-bond formation and sheet-sheet stacking, as a function of peptide concentration $[C_p]$.

Sheet-sheet stacking facilitate amyloidosis. The results in Fig. 3a show that change of $[C_p]$ can reduce ΔG by an extent of $1/4 \sim 1/2$ fold. Also, the stacking itself has a kinetic barrier ΔG_s for a collective onset (i.e., a phase separation between the dilute single-layer and dense multi-layer peptide phases). This ΔG_s is different from ΔG and can be reduced to zero if $[C_p]$ is high enough, Fig. 3b₂. More importantly, in order to have stacking effect, we found that a minimal length of peptide N-terminal repeat Δl_{min} is required, Fig. 3b₁. In other words, the kinetics of non-native β -sheet formations for peptides with $\Delta l < \Delta l_{min}$, is qualitatively similar to the single-layer case. But, once $\Delta l > \Delta l_{min}$, the kinetics can be speeded up by a synergistic inter-sheet stacking. Consequently, non-native β -sheet formations can be more efficient at large Δl and high $[C_p]$ conditions. This finding is consistent with the positive correlation between aggregation tendency and amount of Δl , as observed both clinically and experimentally [3, 5]. Indeed, the minimal number of polyglutamine repeats in Huntington disease is 30-40 [3], in a surprisingly good agreement with our estimates and hence strongly suggesting that *in vivo* aggregation is driven by multi-layer β -sheet stacking.

Template assembly & Nucleated Conversion. The inter-sheet stacking helps β -sheet formation in high peptide concentrations; once formed, these stacked sheets can nevertheless

dissociate if the peptide concentration is down-regulated or fluctuating to a lower level (as a common scene at *in vivo* conditions). But, with a large ΔG , the dissociated sheets can stay at stable single-layer format for a reasonable time span; this would allow them to serve as templates to direct other cytoplasmic peptides into non-native β -sheet format, Fig. 1. Thus, sheet-sheet stacking synergistically enhances non-native β -sheet formation on both thermodynamics and kinetics prospectives. This is a combined nucleated conformational conversion (NCC) and mono/oligomer-directed conversion (M/O-DC) (or template assembly, TA) process [6, 16], and might reconcile the long-standing debate on the self-perpetuating mechanism in amyloid seed formation [6].

The effect of sequestering agent. Finally, we consider the implications of our work for the dose-response curves of therapeutic agents. We consider the simplest case that a scavenger can sequester unfolded mutant peptides before they pass the transition state. In this case, binding of the sequestering agent is *mutually exclusive* with β -sheet H-bond network formation; this is presumably how scavengers such as chaperones or polyamine [7, 19, 20] function. A successful agent, then, should enhance the bottleneck (i.e., minimal mutant peptide concentration $[C_p]_{min}$ and minimal length of repeats Δl_{min}) for the onset of stacking. Using the mean-field approach, we found however that at high sequester concentration, Δl_{min} is reduced (Fig. 3c) because of the modulation of stacking cooperativity (whereas $[C_p]_{min}$ is increased); this suggests that amyloid nucleation is *least* prohibited at median agent concentration. Interestingly, this result seems to correspond with the evidence that huntingtin aggregation is eliminated only via the overexpression or deletion of chaperones [20].

Summary

We have presented a modeling approach to the initial seed formation in protein amyloidosis. Our work indicates that stacking is the critical effect and that simple agents that try to interfere with H-bond network formation may not be very effective at preventing aggregation; alternatively, the kinetics analysis (Fig. 3b) suggests that a potentially more effective strategy would be to attempt to interfere with stacked non-native β -turns or with dissociated non-native single-layer β -sheets. Future work will take into account competition between β -sheets and non-trivial native structures, hence allowing for application of our ideas to a wide variety of aggregation-prone systems. Finally, it is interesting to speculate on the fact that

even given the obvious disadvantage of an aggregation tendency, natural selection has not replaced these oligopeptide-repeat containing proteins by other sequence designs. Perhaps they can serve as an essential building block for bio-architectural construction, [11, 16], or an evolutionary tool which aids in the addition of new sequence to an existing peptide [21], or a buffer to titrate chaperones and hence to expose signaling molecules that have genetic, structural variations buffered by chaperones to environmental challenges [22].

-
- [1] De Young, L.R., Fink, A.L. & Dill, K.A. (1993) *Accounts of Chemical Research* **26**, 614-620.
 - [2] Prusiner, S.B. (1997) *Science* **278**, 245-251.
 - [3] Zoghbi, H.Y. & Orr, H.T. (2000) *Annu. Rev. Neurosci.* **23**, 217-247. (2000).
 - [4] Fändrich, M., Fletcher, M.A. & Dobson, C.M. (2001) *Nature* **410**, 165-166. (2001).
 - [5] Liu, J.J. & Lindquist, S.L. (1999) *Nature* **400**, 573-576.
 - [6] Serio, T.R. *et al.* (2000) *Science* **289**, 1317-1321. (2000).
 - [7] Bercovich, B. *et al.* (1997) *J. Biol. Chem.* **272**, 9002-9010.
 - [8] Ciechanover, A., Orian, A. & Schwartz, A.L. (2000) *Bioessays* **22**, 442-451.
 - [9] Bence, N.F., Sampat, R.M. & Kopito, R.R. (2001) *Science* **292**, 1552-1555.
 - [10] Cheung, M.S., García, A.E., & Onuchic, J.N. (2002) *Proc. Natl. Acad. Sci. USA*. **99**, 685-690.
 - [11] Guo, C., Cheung, M.S., Levine, H. & Kessler, D.A. (2002) *J. Chem. Phys.* **116**, 4353-4365.
 - [12] Guo, C., Levine, H. & Kessler, D.A. in preparation.
 - [13] Wolynes, P.G. (1997) *Proc. Natl. Acad. Sci. USA*. **94**, 6170-6175.
 - [14] Guo, C., Levine, H. & Kessler, D.A. (2000) *Proc. Natl. Acad. Sci. USA*. **97**, 10775-10779.
 - [15] Guo, C., Levine, H. & Kessler, D.A. (2000) *Phys. Rev. Lett.* **84**, 3490-3493.
 - [16] Susan Lindquist, personal communication.
 - [17] Balbirnie, M., Grothe, R. & Eisenberg, D.S. (2001) *Proc. Natl. Acad. Sci. USA*. **98**, 2375-2380.
 - [18] Vargas, R. Garza, J. Dixon, D.A. & Hay, B.P. (2000) *J. Am. Chem. Soc.* **122**, 4750-4755.
 - [19] Supattapone, S. *et al.* (2001) *J. Virol.* **75**, 3453-3461.
 - [20] Krobitsch, S. & Lindquist S. (2000) *Proc. Natl. Acad. Sci. USA*. **97**, 1589-1594.
 - [21] Eckert, R.L. & Green, H. (1986) *Cell* **46**, 583-589.
 - [22] Rutherford, S.L. & Lindquist S. (1998) *Nature* **396**, 336-342.
 - [23] Fersht, A. in *Structure and mechanism in protein science*. (W.H. Freeman & Company. New

York, 1999) Chapter 11.

- [24] Creighton, T.E. in *Proteins: structure and molecular properties* (W.H. Freeman and Company. New York. 1996) Chapter 4 & 5.
- [25] Guo, C. & Levine, H. (1999) *Biophys. J.* **77**, 2358-2365.
- [26] Guo, C. & Levine, H. (2000) *J. Biol. Phys.* **26**, 219-234.

Figure captions

1. The possible kinetics for amyloid nucleation. The N-terminal repeats are sketched by the thin curve connecting the peptide (shadowed) N-, C-terminal domains. The basic choice is whether monomers form aberrant sheets on their own and then either form fibrils directly (monomer spontaneous conversion) or act as templates for further monomer attachment (monomer-directed conversion) or alternatively whether the entire nucleus must form cooperatively (nucleated conformation conversion). Our work suggests a modified picture in which cooperatively formed nuclei may disassociate and act as monomer-directed templates - see text for discussion.
2. **(a)** The phase diagram of a single-layer β -sheet, exhibiting the transition from two-state (region B) to non-two-state (A_1 , A_2) behaviors as a function of sheet topology (M, L). In region A_1 (A_2), there is a dominant intermediate ensemble with a partially folded β -sheet droplet (unfolded bubble) floating along the fibril (hairpin) axis. The inserts show representative specific heat diagrams. **(b)** The predominant kinetic pathway. We computed the free energy change corresponding to a sequence of partially-folded states differing by the addition of one contiguous H-bond and thereby find the transition state and folding barrier for various paths [14, 15]. Analysis shows two predominant patterns: H-bonds are initiated 1) at several β -turns, followed by symmetric propagation (symmetric zipping [14], B_1), or 2) at a collapsed loop and an existing turn, followed by asymmetrically propagation (asymmetric non-zipping, B_2).
3. **(a)** The free energy profile for the predominant kinetics at a given topology and peptide concentration $[C_p]$. ΔG is the free energy barrier height for β -sheet formation within a single peptide and $\delta\Delta G = \Delta G|_{[C_p]} - \Delta G|_{[C_p] \rightarrow 0}$ (inset, dashed curve; note its instability which leads to a sudden jump of $\delta\Delta G$ (solid curve) and hence the onset of stacking). **(b)** The phase diagram including the stacking effect. A minimal

oligopeptide repeat length Δl_{min} is required for stacking (the thin dotted curve and inset **b**₁). Here $\Delta l = (M+1)(L+2)$ (assuming two C_α residues for one β -turn). The kinetics is also modified by the stacking effect (symmetric zipping: C_1 ; asymmetric non-zipping: C_2). **(b**₂) The peptide concentration $[C_p]_{stack}$ at $\Delta l = 90$ required for the onset of stacking and barrier-less stacking; the order of $[C_p]_{stack}$ is compatible with experiment [6]. **(c)** The Δl_{min} and **(c**₁) the minimal peptide concentration $[C_p]_{min}$ (for $\Delta l = 120$) required for barrier-less stacking as a function of agent concentration $[C_{sq}]$. Here $K_d^{coil=N}$ is the dissociation constant between the sequestering agent and a peptide coil that can form a totally N H-bond β -sheet.

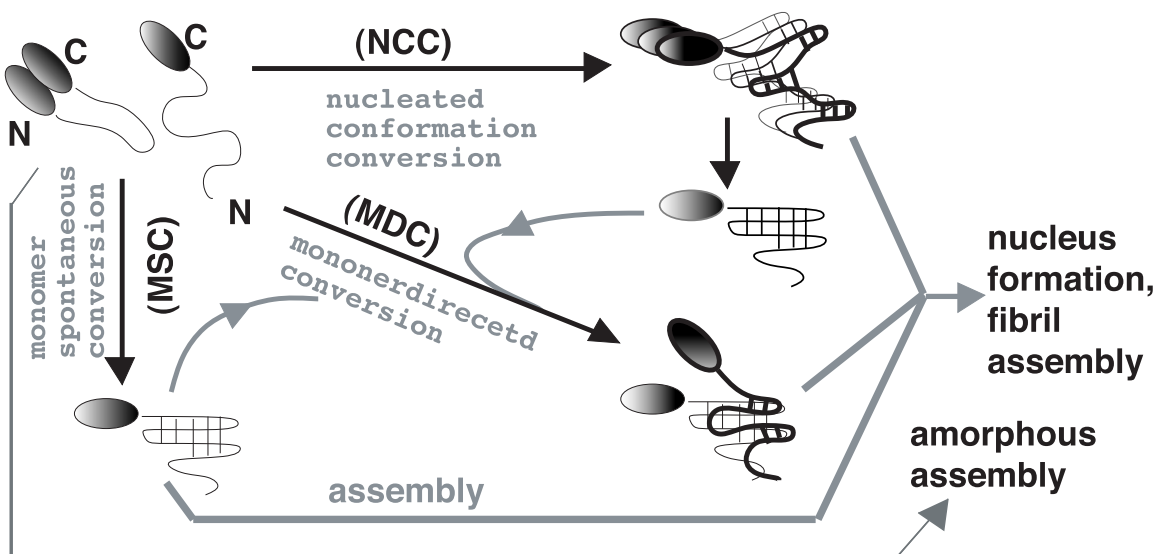


FIG. 1:

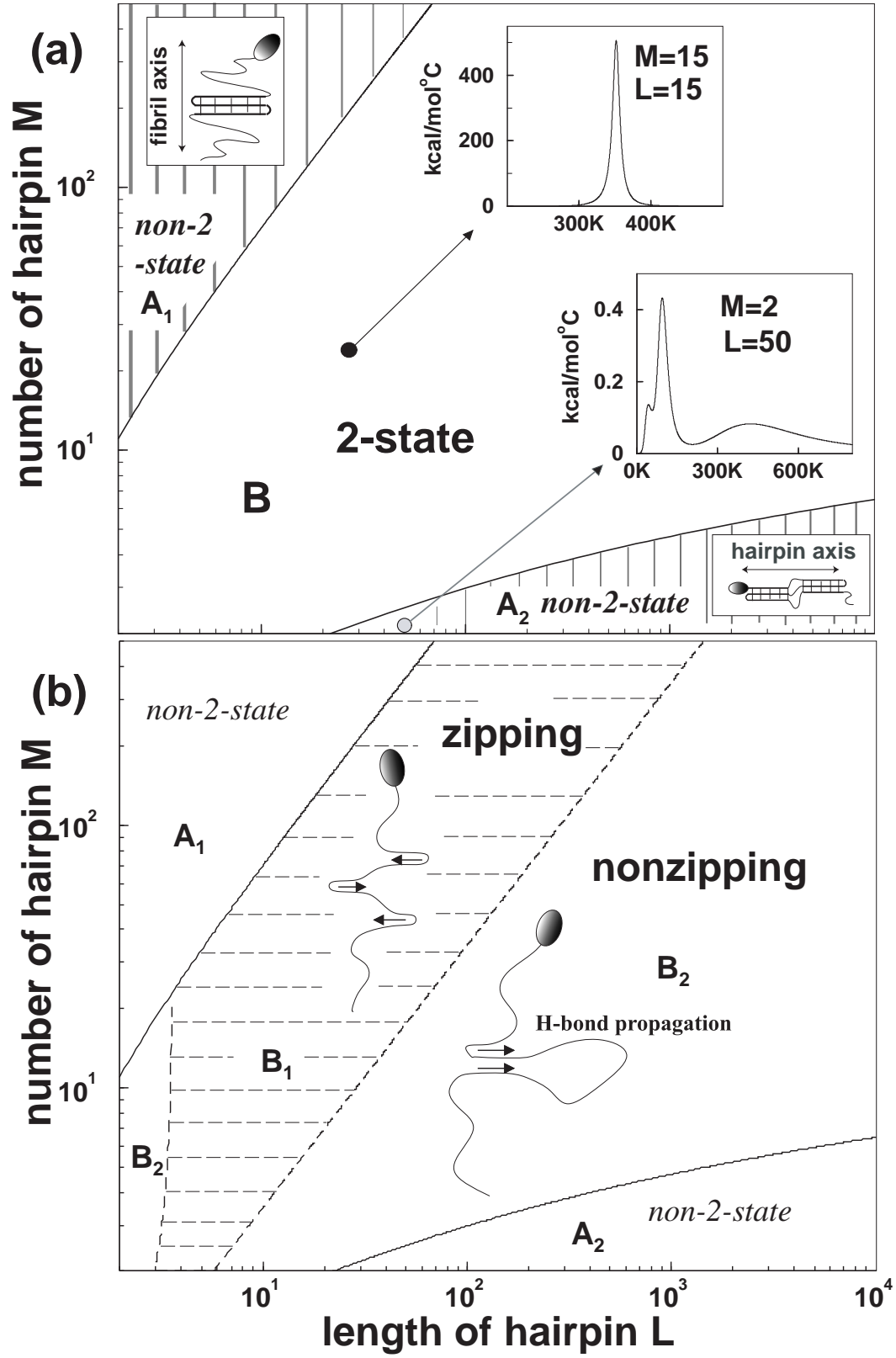


FIG. 2:

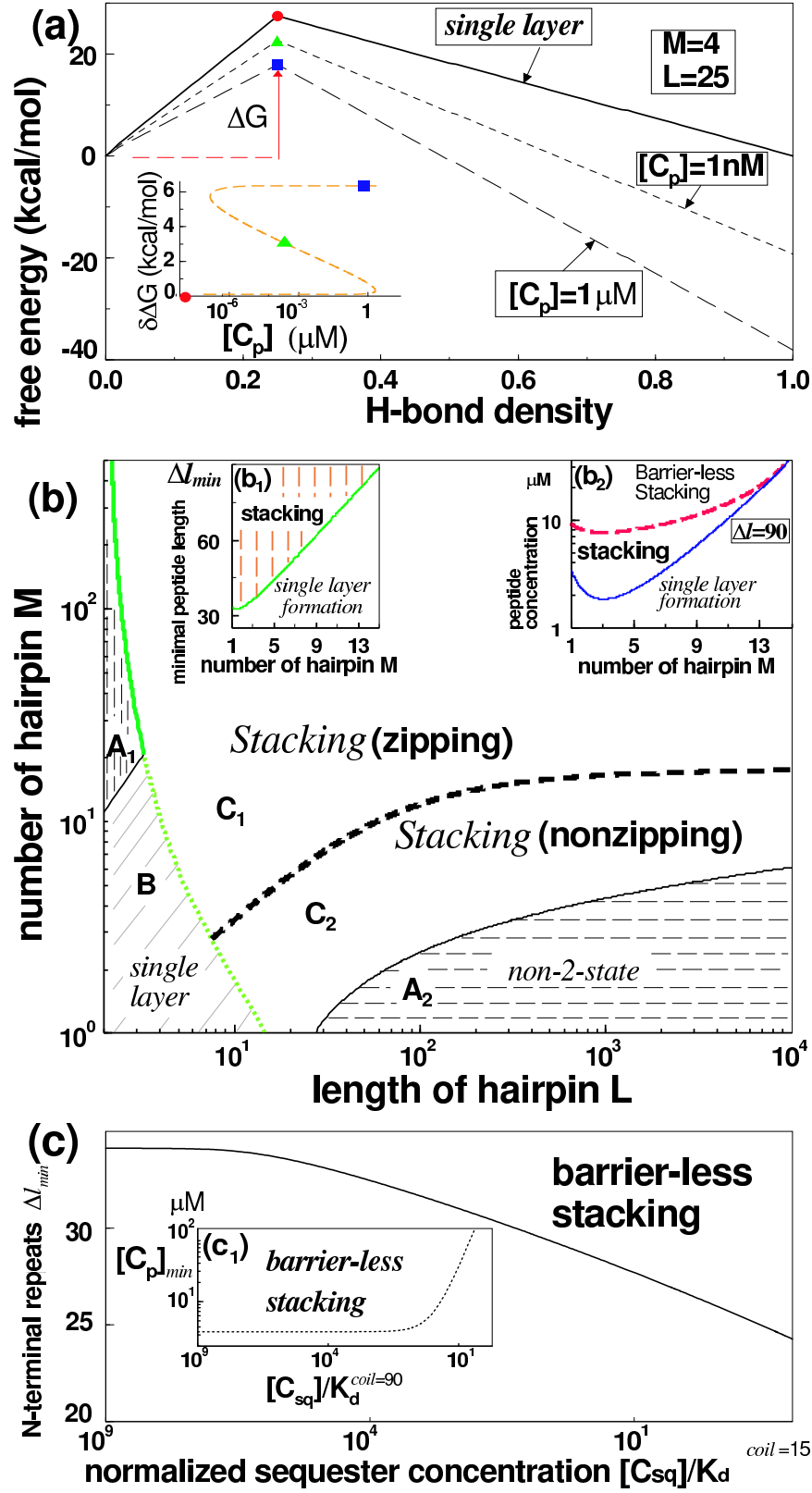


FIG. 3:

Appendix

The Model Hamiltonian

We consider a single isolated sheet consisting of $M + 1$ strands each of length L residues (i.e., number of total H-bonds $N = ML$), with the coordinates of the j^{th} residue ($1 \leq j \leq L$) on the i^{th} -strand ($1 \leq i \leq M$) labeled as $\vec{x}_{i,j}$. Since β -sheet formation primarily involves a competition between the solvation of random coils and the formation of a collective H-bond network, we use an *effective* hydration parameter $\vec{h}_{i,j} \equiv \vec{x}_{i+1,j} - \vec{x}_{i,j}$ to characterize the local H-bonding between adjacent strands. Moreover, as interested in the formation of best β -sheet template for amyloidosis, we monitor the H-bonds between residues labeled by the same index in the two neighboring hairpin strands. Thus, we define a parameter $\Delta_{i,j} = 1$ if $|h_{i,j}| \approx 0$ (i.e., no hydration) and 0 otherwise, to measure the presence of H-bonds correctly contributing to the demanded β -sheet structure. This leads to the following phenomenological model [12]

$$H_{\text{Hb}} = \sum_{i,j} \Delta_{i,j} \left[\frac{1}{2} f_2 \Delta_{i,j \pm 1} + \frac{1}{2} f_3 \Delta_{i \pm 1,j} + f_1 \right] \quad (1)$$

$$H_{\text{coil}} = \frac{1}{2} \kappa \sum_{i,j} \left[1 + \gamma \sum_{k=0}^1 \Delta_{i,j-k} \right] \prod_{k=\pm 1} \left[1 + \gamma_1 \sum_{a=0}^1 \Delta_{i,j-a} \Delta_{i+k,j} \Delta_{i+k,j-1} \right] \left| \vec{h}_{i,j} - \vec{h}_{i,j-1} \right|^2 \quad (2)$$

to describe the energy of correctly formed H-bonds and the entropy of the unfolded coils. Here, f_1 determines the basic energy of formation of a single H-bond and κ the stiffness of an isolated strand. The other parameters are introduced to mimic the cooperativity inherent in H-bond network formation and, as described elsewhere, are fitted by simulations on a more microscopic model [12]. In particular, f_2, γ describe the intra-hairpin coupling along the “hairpin” axis (parallel to the stacked β -strands), and f_3, γ_1 are the inter-hairpin coupling along the “fibril” axis (the fibril direction, perpendicular to hairpin axis).

Of course, when the system becomes larger, there will be mis-paired H-bonds generated between non-adjacent or mis-slid adjacent strands [11]. The lowest order correction to the “bare” partition function (based on Eqns.1,2) is due to one single mis-paired H-bond formed between any two strands. This correction is estimated to be [12]

$$\Gamma_2 \approx \frac{(M+1)M}{2!} \times \frac{4(2-\sqrt{2})}{(1+\gamma)^3} e^{-\beta f_1} L^{1/2} \sqrt{1-\bar{Q}} \stackrel{\text{def}}{=} \frac{(M+1)M}{2!} \sigma_{\bar{Q}} \quad (3)$$

with $\beta = 1/k_B T$, and $\bar{Q} = \sum_{ij} \Delta_{i,j} / N$ is the density of correctly formed H-bonds for the sheet.

Next, to include the stacking effect, we assume that this effect will modify the change of free energy due to the formation of a local, single H-bond, by an amount proportional to the density of ordered H-bonds on nearby sheets. This modification includes both a stacking energy V_s arising from non-specific van der Waals interactions and an entropy loss ΔS due to the reduction of the radius of gyration between partially folded adjacent β -sheets. Details of how we estimated the stacking effect is given in supplement file. Overall, the Hamiltonian for the entire system follows

$$\begin{aligned} H &= \sum_s \{H_{\text{Hb}}(s) + H_{\text{coil}}(s)\} + \frac{1}{2} \sum_{st} J_{st} \sum_{ij} \Delta_{ij,(t)} [g_1 - g_2 \Delta_{ij,(s)}] \\ &\stackrel{\text{def}}{=} \sum_s H_{\text{sing}}(s) + \frac{1}{2} \sum_{st} J_{st} H_{\text{int}}(s, t) \end{aligned} \quad (4)$$

with $J_{st} = 1$ if sheets s, t are stacked together and 0 otherwise, and $\Delta_{ij,(s)}$ is the $\Delta_{i,j}$ parameters for sheet s . Here, we separate the entire Hamiltonian to a single-sheet part H_{sing} and an inter-sheet coupling part H_{int} . Also, for future convenience, we have defined notations $g_1 = \frac{k_B T}{N} \ln \left[\frac{\Omega_0}{\Omega_N} \right]$ and $g_2 = N^{-1}[-V_s + g_1]$, where Ω_0 and Ω_N are the volume of completely unfolded coil and folded β -sheet, respectively.

The Partition Function

Without Stacking Effect. Given the above expression, we transform the β -sheet into a system with many coupled hairpins. Then, for a given choice of topology (M, L) and contact configuration $\{\Delta_{i,j}\}$ (which then yields a particular H-bond density \bar{Q}), we found that the partition function can be dissected in terms of three parts

$$Z[\{\Delta_{i,j}\}] = Z_{\text{Hb}} Z_{\text{coil}} [1 + \Gamma_2] \quad (5)$$

that allows us to compute contributions from the entropy of coiled part (in Z_{coil}), the energy of partially formed H-bond (in Z_{Hb}), and the H-bond mis-pairing effect (in Γ_2). Again, the details of derivation can be found in supplement file.

For the thermodynamics, we are interested in whether at the folding temperature T_f (at which the partition function $Z_{\bar{Q}}$ of a random coil, $\bar{Q} = 0$, and the completely folded state, $\bar{Q} = 1$, are equally weighted), there are other competing intermediates $Z_{\bar{Q}_{\text{int}}}$. In principle, for a two-dimensional β -sheet structure, there are two possible competing intermediates with a partially folded droplet (unfolded bubble) floating along the fibril (hairpin) axis with

translational entropy large enough to compete with the completely un/folded states Z_0 , Z_1 . The phase boundaries separating the dominance of these ensemble are illustrated in Fig. 2 with representative specific heat diagrams inserted. Our approach to the study of the kinetics of folding is described in the caption to Fig. 2 and elsewhere [14, 15].

With Stacking Effect. Using variational mean field theory to approach the thermodynamics, we obtained the following self-consistent equation

$$\langle \bar{Q} \rangle = \frac{e^{\beta\mu} \sum_{\bar{Q}} \bar{Q} Z_{\bar{Q}} [\langle \bar{Q} \rangle]}{1 + e^{\beta\mu} \sum_{\bar{Q}} Z_{\bar{Q}} [\langle \bar{Q} \rangle]} \quad (6)$$

$$Z_{\bar{Q}} [\langle \bar{Q} \rangle] = Z_{\bar{Q}}[0] e^{-N[2\beta V_s \bar{Q} + N^{-1}(1-2\bar{Q}) \ln \frac{\Omega_0}{\Omega_N}]} \langle \bar{Q} \rangle \stackrel{\text{def}}{=} Z_{\bar{Q}}[0] e^{h_1 \langle \bar{Q} \rangle + h_2 \bar{Q} \langle \bar{Q} \rangle} \quad (7)$$

where $Z_{\bar{Q}}[0]$ is the bare partition function without the stacking effect and μ is the peptide chemical potential. Apparently, μ is related to the peptide concentration $[C_p]$ and the relation is estimated as

$$[C_p] \Omega_0 \approx \left. \frac{e^{\beta\mu} Z_0[\bar{Q}]}{1 + e^{\beta\mu} Z_0[\bar{Q}]} \right|_{\langle \bar{Q} \rangle \approx 0} \quad (8)$$

To identify where the two-state-like behavior emerges, we examined if any of the aforementioned intermediates dominates at the temperature T_f where $Z_0[0] = Z_1[0]$. Apparently, if \bar{Q}_{int} is the H-bond density of the dominant ensemble, we would have $\langle \bar{Q} \rangle = \bar{Q}_{int}$. This implies that the two-state behavior exists if $\forall \bar{Q}_{int}$, we have $(0 < \bar{Q}_{int} < 1)$

$$Z_{\bar{Q}_{int}}[\bar{Q}_{int}] = Z_{\bar{Q}_{int}}[0] e^{(h_1 + h_2 \bar{Q}_{int}) \bar{Q}_{int}} < \max [Z_0[\bar{Q}_{int}], Z_1[\bar{Q}_{int}]] \quad (9)$$

for a given topology (M, L) and temperature T_f . Solving this requirement, we obtained the phase diagram Fig. 3a; likewise, the predominant kinetic patterns can be computed in the 2-state region (see supplement for details).

Estimating the Minimal Length for Stacking

At the two-state region, the ensemble sum in (6) can be reduced as $\sum_{\bar{Q}} Z_{\bar{Q}} [\langle \bar{Q} \rangle] \approx Z_0[\langle \bar{Q} \rangle] + Z_1[\langle \bar{Q} \rangle]$; thus after re-arrangement, Eqn. (6,7) can be re-formulated into

$$[C_p] \Omega_0 = \frac{\langle \bar{Q} \rangle}{1 - \langle \bar{Q} \rangle} e^{-f_{stack} \langle \bar{Q} \rangle} \quad (10)$$

where $f_{stack} = 2[N\beta V_s + \ln(\Omega_0/\Omega_N)]$. Then, from a previous study [25, 26], we realize that Eqn. (10) can yield a “van-der-Waals” loop in the $([C_p], \langle \bar{Q} \rangle)$ diagram and hence a phase

separation (coexistence) effect. The phase separation referred to here is a separation between the dilute random coiled phase and the dense stacking phase. Numerical calculation showed that this requires a minimal length Δl_{min} , as shown in Fig. 3a₁. The peptide concentration that allows for phase separation can also be computed (see supplement for details); this yields Fig. 3a₂.

Effect of Sequesterers.

The effect of sequesterers on the β -sheet partition function is estimated as the following

$$Z_{\bar{Q}} [\langle \bar{Q} \rangle] \rightarrow Z_{\bar{Q}} [\langle \bar{Q} \rangle] \left[1 + \frac{(1 - \bar{Q}') [C_{sq}]}{K_d [\{\Delta_{ij}\}]} \right] \quad (11)$$

where $[C_{sq}]$ is the sequesterer concentration, $K_d [\{\Delta_{ij}\}]$ is the β -sheet configuration-dependent dissociation constant with the sequesterer, \bar{Q}' is the averaged H-bond density from other sheets nearby the one interacting with sequesterers. Here the prefactor $(1 - \bar{Q}')$ indicates that binding of sequesterer is unlikely to occur if the target peptide is tightly surrounded by other sheets (as they tend to stack together into an anhydrous, dense aggregate).

Mutually exclusive effects. One sequesterer might generally bind to multiple amide (or carbonyl) groups on a single targeted peptide; here for simplicity, we assume that the binding of each individual sequesterer functional site to the peptide amide (or carbonyl) group occurs in an independent manner. Moreover, we assume that the binding is *mutually exclusive* with local H-bond formation, as presumably how scavengers such as polyamine [19] function. Then, for a pathological peptide that can form N H-bonds in total, assuming K_d^{coil} as the dissociation constant for a fully coiled peptide, we found that the peptide concentration for sheet-sheet stacking is modified (see supplement for detailed derivation)

$$[C_p] \Omega_0 = \left[1 + \frac{[C_{sq}]}{K_d^{coil}} (1 - \langle \bar{Q} \rangle) \right] \frac{\langle \bar{Q} \rangle}{1 - \langle \bar{Q} \rangle} e^{-\langle \bar{Q} \rangle f_{stack}} \quad (12)$$

Indeed, we found that the cooperativity in sheet-sheet stacking is enhanced by the presence of sequesterers. This is because the binding of sequesterer and β -sheet H-bond formation as well as stacking are *mutually exclusive*. Thus, the thermodynamic weighting of intermediates with partially formed H-bonds and weakly stacked β -sheets are significantly suppressed; only completely folded and well-stacked sheets can escape from the attack of sequesterers. In the presence of sequesterer, we are interested in its dose-response curve. Specifically, we compute the minimal peptide concentration $[C_p]_{min}$ as well as minimal length

of repeats Δl_{min} required for the downhill stacking. The numerical results are shown in Fig. 3b.

Supporting information for “Non-native β -sheet formation: insights into protein amyloidosis”

Chinlin Guo^{†‡}, Herbert Levine[‡], and David A. Kessler^{*}

*[†]Department of Molecular Cell Biology,
Harvard University, 16 Divinity Avenue,
Room 3007, Cambridge, MA 02138*

*[‡]Department of Physics, University of California,
San Diego 9500 Gilman Drive, La Jolla, CA 92093-0319 and
^{*}Department of Physics, Bar-Ilan University, Ramat-Gan, Israel*

I. ESTIMATING THE STACKING EFFECT

To proceed, we first notice that V_s is actually sequence-dependent. In general, energies of this type range from -10^{-2} to -0.2 kcal/mol [1]. As we are interested in the generic behavior independent of sequence, we take $V_s = -0.1$ kcal/mol as the size of the non-specific C_α - C_α coupling [1] to mimic the stacking energy. Since this non-specific interaction between two H-bond units (from two different, nearby sheets) occurs only when both units have formed H-bonds, it contributes an energy between two stacked sheet s and t : $H_{V_s}(s, t) = \sum_{i,j} \Delta_{ij,(s)} \Delta_{ij,(t)} V_s$, where $\Delta_{ij,(s)}$ ($\Delta_{ij,(t)}$) is for the $(i, j)^{\text{th}}$ H-bond at sheet s (t). The stacking energy for the entire system then follows $E_{V_s} = \frac{1}{2} \sum_{st} J_{st} H_{V_s}(s, t)$ where $J_{st} = 1$ if s, t are stacked sheets and 0 otherwise. Note that there are at most two sheet t 's that can stack with sheet s .

Second, we notice that the entropy reduction in stacked sheets arises merely from their structural conflicts. When the sheets are all in their completely folded state, there is no entropy reduction upon stacking since they are already in their minimal entropy state. Similarly, wandering of random coils is less prohibited when contiguous to a completely unfolded sheet, compared to a partially or completely folded one. For simplicity, we estimate the entropy of different conformations based on their volume. As an example, the volume of a fully unfolded coil, Ω_0 , can be estimated from the random-flight chain model [2] $\Omega_0 \approx [\sqrt{l} R_{C_\alpha}]^3$ where l is the number of total residues at the mutant peptide N-terminal repeats, and the unit length for radius of gyration $R_{C_\alpha} \approx 11.4 \text{\AA}$ [2]. Likewise, the volume Ω_N of a completely folded β -sheet with a total of N H-bonds ($N = ML$) can be estimated as $\Omega_N = N v_{\text{Hb}}$ where $v_{\text{Hb}} \approx 4.8 \times 3.8 \times 10.0 \text{\AA}^3$ is the unit volume of a single H-bond measured in a densely-packed β -sheet aggregate [3].

Now, to estimate the entropy reduction and keep the formula simple, we use a global term instead of counting all local effects as done in the case of V_s . Specifically, for a given sheet s that is surrounded by two other sheets, indexed by t , we introduce the quantity $\bar{Q}' = \sum_t \sum_{ij} \Delta_{ij,(t)} / (2N)$ to account for the averaged H-bond density in these two sheets, as a global indication of how well they form β -sheet structures. Moreover, for simplicity we assume that the entropy loss of sheet s can be roughly accounted for by a progressive volume reduction from Ω_0 to Ω_N due to the structural conflict with its stacking neighbors, and for which we take the simple form $H_{\Delta S}(s) = (1 - \bar{Q}) \bar{Q}' k_B T \ln[\Omega_0 / \Omega_N]$. Here, the factor $(1 - \bar{Q}) \bar{Q}'$

indicate the structural conflict between sheet s and its contiguous neighbors. Of course, one can devise a more complicated formula to account for how the presence of H-bonds affects the entropy reduction, but this simple form will suffice for our present needs.

Overall, the Hamiltonian for the entire system follows

$$\begin{aligned}
H &= \sum_s [H_{\text{Hb}}(s) + H_{\text{coil}}(s)] + E_{V_s} + \frac{k_B T}{2N} \ln \left[\frac{\Omega_0}{\Omega_N} \right] \sum_{st} J_{st} \Delta_{ij,(t)} \left[1 - \frac{\Delta_{ij,(s)}}{N} \right] \\
&\stackrel{\text{def}}{=} \sum_s \{H_{\text{Hb}}(s) + H_{\text{coil}}(s)\} + \frac{1}{2} \sum_{st} J_{st} \sum_{ij} \Delta_{ij,(t)} [g_1 - g_2 \Delta_{ij,(s)}] \\
&\stackrel{\text{def}}{=} \sum_s H_{\text{sing}}(s) + \frac{1}{2} \sum_{st} J_{st} H_{\text{int}}(s, t)
\end{aligned} \tag{1}$$

with the notation defined in main text.

II. ESTIMATING THE PARTITION FUNCTION

A. Without Stacking Effect

Given the Hamiltonian

$$H_{\text{Hb}} = \sum_{i,j} \Delta_{i,j} \left[\frac{1}{2} f_2 \Delta_{i,j \pm 1} + \frac{1}{2} f_3 \Delta_{i \pm 1,j} + f_1 \right] \tag{2}$$

$$H_{\text{coil}} = \frac{1}{2} \kappa \sum_{i,j} \left[1 + \gamma \sum_{k=0}^1 \Delta_{i,j-k} \right] \prod_{k=\pm 1} \left[1 + \gamma_1 \sum_{a=0}^1 \Delta_{i,j-a} \Delta_{i+k,j} \Delta_{i+k,j-1} \right] \left| \vec{h}_{i,j} - \vec{h}_{i,j-1} \right|^2, \tag{3}$$

we can transform the β -sheet into a system with many coupled hairpins. Then, for a given choice of topology (M, L) and contact configuration $\{\Delta_{i,j}\}$ (which then yields a particular H-bond density \bar{Q}), we further dissected each of these hairpins into one or several unfolded coils (“bubbles”) separated by successively folded segments (“droplets”). The partition function for these unfolded coils is estimated by previously developed methods [6, 7]. Specifically, we define a functional $M_i(j_1, j_2)$ as the Gaussian path integral (based on Eqn.3) for a coil running from residue index $(i, j_1 + 1)$ to $(i, j_2 - 1)$ with both (i, j_1) , (i, j_2) being H-bonded and a functional $W_i(j_1, j_2)$ for a continuously H-bonded segment from index (i, j_1) to (i, j_2) (based on Eqn.2)

$$M_i(j_1, j_2) = \prod_{j=j_1+1}^{j_2-1} \int \left[\frac{\beta \kappa}{2\pi} \right]^{\frac{3}{2}} d\vec{h}_{i,j} (1 - \Delta_{i,j})$$

$$\times e^{-\beta H_{\text{Hb}}(\{\Delta_{i,j}\})} \quad (4)$$

$$W_i(j_1, j_2) = \prod_{j=j_1}^{j_2} \int \left[\frac{\beta \kappa}{2\pi} \right]^{\frac{3}{2}} d\vec{h}_{i,j} \Delta_{i,j} e^{-\beta H_{\text{coil}}(\{\Delta_{i,j}\})} \quad (5)$$

These Gaussian integrals can be worked out as shown in Ref. [6, 7, 8]. The partition function can be broken up into successive multiplication of these functionals

$$\begin{aligned} Z[\{\Delta_{i,j}\}] &= \prod_{i,j} \int \left[\frac{\beta \kappa}{2\pi} \right]^{\frac{3}{2}} d\vec{h}_{i,j} [(1 - \Delta_{i,j}) + \Delta_{i,j}] e^{-\beta[H_{\text{Hb}}(\{\Delta_{i,j}\}) + H_{\text{coil}}(\{\Delta_{i,j}\})]} [1 + \Gamma_2(\bar{Q})] \\ &= \left\{ \prod_i [W_i(0, j_1) M_i(j_1, j_2) W_i(j_2, j_3) \cdots] \right\} [1 + \Gamma_2(\bar{Q})] \stackrel{\text{def}}{=} Z_{\text{Hb}} Z_{\text{coil}} [1 + \Gamma_2] \end{aligned} \quad (6)$$

where $(i, j_1), (i, j_2), \dots$ are the end points for the successive folded segments. Thus, in this way, we can dissect the free energy in terms of contributions from the Gaussian integral (in Z_{coil}), from the *Ising-model-like* energy component (in Z_{Hb}), and from the H-bond mispairing effect (in Γ_2). This allows us to fit the parameters with simulations on a microscopic model [8]; once obtained, these parameters can be used to retrieve the entire partition function and hence the density of states using Multicanonical Monte Carlo sampling [9].

B. With Stacking Effect

In the presence of stacking effect, the partition functional for the entire system reads

$$\mathcal{Z} = \int \mathcal{D}h e^{-\beta \sum_s [H_{\text{sing}}(s) + \frac{1}{2} \sum_t J_{st} H_{\text{int}}(s, t)]} \quad (7)$$

where $\int \mathcal{D}h$ indicates integration for the entire $\vec{h}_{i,j}(s)$ vector space. To compute the thermodynamics, we use a Gaussian trick implemented in Ref. [4, 5] to decouple one of the inter-sheet terms. Briefly, we insert a Gaussian integral identify in Eqn.(7) to separate the $\Delta_{ij,(s)} \Delta_{ij,(t)}$ terms

$$\begin{aligned} \mathcal{Z} &\rightarrow \int \mathcal{D}h e^{-\beta \sum_s [H_{\text{sing}}(s) + \frac{g_1}{2} \sum_t J_{st} \Delta_{ij,(t)}]} \int \mathcal{D}\zeta e^{-\frac{1}{2\beta g_2} \sum_{st} J_{st}^{-1} \sum_{ij} \zeta_{ij,(s)} \zeta_{ij,(t)}} \\ &\quad \times e^{\sum_s \Delta_{ij,(s)} \zeta_{ij,(s)}} \end{aligned} \quad (8)$$

Then, using a transform $\zeta_{ij,(s)} = \beta g_2 \sum_t J_{st} \eta_{ij,(t)}$ and the identity $\sum_t J_{st} \Delta_{ij,(t)} = 2N\bar{Q}'(s)$, we have

$$\begin{aligned} \mathcal{Z} &\rightarrow \int \mathcal{D}\eta e^{-\frac{\beta g_2}{2} \sum_{st} J_{st} \sum_{ij} \eta_{ij,(s)} \eta_{ij,(t)}} \int \mathcal{D}h e^{-\beta \sum_s [H_{\text{sing}}(s) + N g_1 \bar{Q}'(s)]} \\ &\quad \times e^{\beta g_2 \sum_{st} J_{st} \sum_{ij} \Delta_{ij,(s)} \eta_{ij,(t)}} \end{aligned} \quad (9)$$

Next, we use variational mean field theory to approach the thermodynamics. First, we set $\bar{Q}' = \langle \bar{Q} \rangle$ in Eqn.(9) as the thermodynamically averaged H-bond density. Second, we rewrite the partition function into a form of grand canonical ensemble by inserting a chemical potential μ for each peptide. These steps allow us to decouple the partition function Eqn.(9) into a product of independent single peptide partition function where each peptide interacts with a background field $\{\eta\}$. Finally, the connection between $\{\eta\}$ and $\langle \bar{Q} \rangle$ can be found by a variational approach on $\{\eta\}$ similar to the procedure used in Ref. [4, 5]. This yields a self-consistent relation $\eta_{ij,(s)} = \langle \Delta_{ij,(s)} \rangle$ where $\langle \Delta_{ij,(s)} \rangle$ is the thermodynamic expectation value, which by definition is equivalent to $\langle \bar{Q} \rangle$ in the sense of mean-field approach. Overall, one can work out the following self-consistent equation

$$\langle \bar{Q} \rangle = \frac{e^{\beta\mu} \sum_{\bar{Q}} \bar{Q} Z_{\bar{Q}} [\langle \bar{Q} \rangle]}{1 + e^{\beta\mu} \sum_{\bar{Q}} Z_{\bar{Q}} [\langle \bar{Q} \rangle]} \quad (10)$$

$$\begin{aligned} Z_{\bar{Q}} [\langle \bar{Q} \rangle] &= Z_{\bar{Q}}[0] e^{\beta[-g_1 + 2g_2 \sum_{ij} \Delta_{i,j}] \langle \bar{Q} \rangle} \\ &= Z_{\bar{Q}}[0] e^{-N \left[2\beta V_s \bar{Q} + N^{-1} (1-2\bar{Q}) \ln \frac{\Omega_0}{\Omega_N} \right] \langle \bar{Q} \rangle} \\ &\stackrel{\text{def}}{=} Z_{\bar{Q}}[0] e^{h_1 \langle \bar{Q} \rangle + h_2 \bar{Q} \langle \bar{Q} \rangle} \end{aligned} \quad (11)$$

where $Z_{\bar{Q}}[0]$ is defined as in the main text. Note that the entropy effect is scaled by N^{-1} and hence is less important than the energy term.

The loss of translational entropy of one peptide due to the presences of other peptides is absorbed into the chemical potential μ in this mean-field approach [4]. To estimate the relation between μ and peptide concentration $[C_p]$, we use a dilute gas of weakly folded peptides as the basic state which then “competes” with aggregate formation. In other words, the peptide concentration is that before the onset of aggregation. Specifically, we estimated the probability of finding a partially folded peptide in a unit volume that can be caught by other peptides. The size of this volume is estimated roughly to $V_{\bar{Q}} \approx 4\pi[R_G(\bar{Q})]^3/3$ with $R_G(\bar{Q})$ and \bar{Q} denoting the radius of gyration and H-bond density of that peptide, respectively. $R_G(\bar{Q})$ is further estimated by a linear interpolation $R_G(\bar{Q}) = (1 - \bar{Q})\sqrt{l}R_{C_\alpha} + \bar{Q}[v_{\text{Hb}}]^{1/3}$ between the volume estimates of fully un/folded β -sheets. Then, $[C_p]$ is connected to μ via the probability P_p of finding a peptide in volume $V_{\bar{Q} \approx 0}$

$$[C_p]V_{\bar{Q} \approx 0} \approx \frac{1}{\beta} \frac{\partial}{\partial \mu} \ln \left[1 + e^{\beta\mu} \sum_{\bar{Q} \ll 1} Z_{\bar{Q}}[\langle \bar{Q} \rangle] \right]_{\langle \bar{Q} \rangle \approx 0}$$

$$\approx \frac{e^{\beta\mu} Z_0[\bar{Q}]}{1 + e^{\beta\mu} Z_0[\bar{Q}]} \Big|_{\langle \bar{Q} \rangle \approx 0} \quad (12)$$

where $V_{\bar{Q} \approx 0} \approx \Omega_0$. Thus, for any given M, L, and T, we can compute $Z_{\bar{Q}}[0]$ by Monte Carlo simulation, and then combine the given $[C_p]$ to solve $\langle \bar{Q} \rangle$ self-consistently.

III. MINIMAL LENGTH FOR STACKING

At the two-state region, the ensemble sum in Eqn. (10) can be reduced as $\sum_{\bar{Q}} Z_{\bar{Q}}[\langle \bar{Q} \rangle] \approx Z_0[\langle \bar{Q} \rangle] + Z_1[\langle \bar{Q} \rangle]$; thus after re-arrangement, Eqn. (10,11) can be re-formulated by defining an effective chemical potential μ'

$$e^{\beta\mu'} \stackrel{\text{def}}{=} \frac{Z_1[0] e^{\beta\mu - \langle \bar{Q} \rangle \ln \frac{\Omega_0}{\Omega_N}}}{1 + Z_0[0] e^{\beta\mu - \langle \bar{Q} \rangle \ln \frac{\Omega_0}{\Omega_N}}} \quad (13)$$

$$\begin{aligned} \langle \bar{Q} \rangle &= \frac{e^{\beta\mu} Z_1[\langle \bar{Q} \rangle]}{1 + e^{\beta\mu} (Z_0[\langle \bar{Q} \rangle] + Z_1[\langle \bar{Q} \rangle])} \\ &= \frac{e^{\beta\mu'} e^{2[N\beta V_s + \ln \frac{\Omega_0}{\Omega_N}] \langle \bar{Q} \rangle}}{1 + e^{\beta\mu'} e^{2[N\beta V_s + \ln \frac{\Omega_0}{\Omega_N}] \langle \bar{Q} \rangle}} \\ &\stackrel{\text{def}}{=} \frac{e^{\beta\mu' + f_{stack} \langle \bar{Q} \rangle}}{1 + e^{\beta\mu' + f_{stack} \langle \bar{Q} \rangle}} \end{aligned} \quad (14)$$

$$\Rightarrow \beta\mu' = -f_{stack} \langle \bar{Q} \rangle - \ln \left[\frac{1}{\langle \bar{Q} \rangle} - 1 \right] \quad (15)$$

where we have used the condition that the temperature is set at $Z_0[0] = Z_1[0]$. Then, from Eqn. (12,13) we have

$$\begin{aligned} [C_p] V_{\bar{Q} \approx 0} &\approx \frac{e^{\beta\mu} Z_0[\bar{Q}]}{1 + e^{\beta\mu} Z_0[\bar{Q}]} \Big|_{\langle \bar{Q} \rangle \approx 0} \\ &= \frac{Z_0[0] e^{\beta\mu - \langle \bar{Q} \rangle \ln \frac{\Omega_0}{\Omega_N}}}{1 + Z_0[0] e^{\beta\mu - \langle \bar{Q} \rangle \ln \frac{\Omega_0}{\Omega_N}}} \equiv e^{\beta\mu'} \end{aligned} \quad (16)$$

After a straightforward calculation, one can show that the van-der-Waals in $(\mu', \langle \bar{Q} \rangle)$ diagram occurs only if $f_{stack} > 4$. Numerically, one can show that f_{stack} has a dependence on the β -sheet topology as well as the mutant polymer length Δl (the N-terminal oligopeptide repeats that construct the non-native β -sheet). Thus, the criteria $f_{stack} > 4$ yields a minimal length Δl_{min} requirement for the onset of stacking effect, as shown in Fig. 3a₁. Also, from the abovementioned study [4, 5], we note that Eqn. (15) has one spinodal point at $\langle \bar{Q} \rangle_{s-} = [1 - \sqrt{1 - 4/f_{stack}}]/2$ where the barrier for the onset of stacking is zero (downhill

stacking), and one binodal point $\langle \bar{Q} \rangle_{b-} = [1 - \xi]/2$ with $2\xi = \tanh[f_{stack}\xi/2]$ where the dilute and dense phases are equally weighted. The corresponding peptide concentrations $[C_p]$ for these two points can be computed from Eqns. (13,15,16)

$$[C_p]V_{\langle \bar{Q} \rangle} = \frac{\langle \bar{Q} \rangle}{1 - \langle \bar{Q} \rangle} e^{-\langle \bar{Q} \rangle f_{stack}}, \quad (17)$$

yielding Fig. 3a₂.

IV. EFFECT OF SEQUESTERERS

To model the effect of sequesterers, we note that the partition function for one protein A when interacting with another one B, $A+B \leftrightarrow AB$, has a general form $Z_A = Z_A[0] + \frac{[C_B]}{K_d} Z_A[1]$ where $Z_A[0]$, $Z_A[1]$ are the partition function of protein A in the absence, presence of protein B, respectively, and K_d is the dissociation constant of the reaction. More generally, the second term can be expressed as the sum in which various K_d 's, each representing a different configuration of protein A, are included. Then, examining the effect of one sequesterer interacting with one non-native β -sheet, we can see that it modifies the partition function eqn.(6) as follows:

$$Z_{\bar{Q}}[\langle \bar{Q} \rangle] \rightarrow Z_{\bar{Q}}[\langle \bar{Q} \rangle] \left[1 + \frac{(1 - \bar{Q}') [C_{sq}]}{K_d [\{\Delta_{ij}\}]} \right] \quad (18)$$

To express the form of $K_d [\{\Delta_{ij}\}]$, we further assume that the binding of each individual sequesterer functional site to the peptide amide (or carbonyl) group occurs in an independent manner. This leads to a multiplicative expansion of the second term in eqn.(18)

$$\frac{[C_{sq}]}{K_d [\{\Delta_{ij}\}]} = e^{\beta \mu_{sq}} \left\{ \prod_{i,j} [1 - \Delta_{ij}] [1 + e^{\beta V_{sq}}] - 1 \right\} \quad (19)$$

Here, μ_{sq} is the sequesterer chemical potential. The factor $[1 - \Delta_{ij}]$ indicates that a local interaction between sequesterer and the peptide occurs only if the local H-bond is not formed, which then allows a local binding or unbinding via a binding energy V_{sq} roughly estimated to be -2.8 kcal/mol comparable with the order of the H-bond energy (which is presumably how sequesterer binds to the amide or carbonyl group). Finally, the term “-1” accounts for the complete non-binding event and ensures at least one binding between the sequesterer and the peptide. In general, one can include more complicated binding pattern but here we restrict our modeling to this simplest case.

Then, for a mutant peptide that can form N H-bonds in total, we can rewrite eqn.(13,15,17) as

$$\beta\mu' = -f_{stack}\langle\bar{Q}\rangle + \ln \left[1 + \frac{[C_{sq}]}{K_d^{coil}} (1 - \langle\bar{Q}\rangle) \right] - \ln \left[\frac{1}{\langle\bar{Q}\rangle} - 1 \right] \quad (20)$$

$$[C_p]V_{\langle\bar{Q}\rangle} = \left[1 + \frac{[C_{sq}]}{K_d^{coil}} (1 - \langle\bar{Q}\rangle) \right] \frac{\langle\bar{Q}\rangle}{1 - \langle\bar{Q}\rangle} e^{-\langle\bar{Q}\rangle f_{stack}} \quad (21)$$

-
- [1] Fersht, A. in *Structure and mechanism in protein science*. (W.H. Freeman & Company. New York, 1999) Chapter 11.
 - [2] Creighton, T.E. in *Proteins: structure and molecular properties* (W.H. Freeman and Company. New York. 1996) Chapter 4 & 5.
 - [3] Balbirnie, M., Grothe, R. & Eisenberg, D.S. (2001) *Proc. Natl. Acad. Sci. USA*. **98**, 2375-2380.
 - [4] Guo, C. & Levine, H. (1999) *Biophys. J.* **77**, 2358-2365.
 - [5] Guo, C. & Levine, H. (2000) *J. Biol. Phys.* **26**, 219-234.
 - [6] Guo, C., Levine, H. & Kessler, D.A. (2000) *Proc. Natl. Acad. Sci. USA*. **97**, 10775-10779.
 - [7] Guo, C., Levine, H. & Kessler, D.A. (2000) *Phys. Rev. Lett.* **84**, 3490-3493.
 - [8] Guo, C., Levine, H. & Kessler, D.A. in preparation.
 - [9] Hansmann, U.H.E, Okamoto, Y. & Eisenmenger, F. (1996) *Chem. Phys. Lett.* **259**, 321-330.